been included in the Brief Description of the Drawings. The present amendment adds no new matter as the sequences are already disclosed in the application as filed.

Applicants further submit a substitute paper and computer readable format (CRF) copy of the sequence listing. The substitute paper and CRF sequence listings add no new matter to the present application, but merely serve to identify sequences already present in the application as filed, and to correct the discrepancies in serial number, filing date, and priority data noted by the Examiner in her Communication.

Summary

Applicants respectfully submit that the present amendment puts the application in condition for allowance. Allowance of the application is respectfully requested at the earliest possible date.

Respectfully submitted,

KAZAZIAN, JR. ET AL.

July 30, 2002
(Date)

By: Kathryn Doyle

KATHRYN DOYLE, Ph.D, J.D.
Registration No. 36,317
MORGAN, LEWIS & BOCKIUS, LLP
1701 Market Street
Philadelphia, PA 19103-2921
Telephone: (215) 963-5000
Direct Dial: (215) 963-4723
Facsimile: (215) 963-5299
E-Mail: kdoyle@morganlewis.com
Attorney for Applicants

KD/JDGB lcf

Enclosures: Marked up copy of the "Brief Description of the Drawings"
Clean copy of the "Brief Description of the Drawings"
Substitute CRF copy of the sequence listing
Substitute paper copy of the sequence listing
Statement pursuant to 37 C.F.R.§1.821-1.825

Marked-Up Copy of the Brief Description of the Drawings

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a diagram depicting the organization of a 6.0 kb human L1 element. ORF1 and ORF2 are indicated by dark rectangles; the 5' and 3' untranslated regions are indicated by shaded rectangles and the untranslated region between ORF1 and ORF2 is indicated by a white stripe. The approximate position of the endonuclease (EN), reverse transcriptase (RT), cysteine-rich C. motif and poly A tail (AAAAA)n are indicated. Arrows indicate the target site duplications which flank the element.

Figure 1B is a diagram of an overview of a retrotransposition assay. The element L1.2 was tagged with an indicator gene (*mneoI*) containing an antisense copy of the *neo* gene disrupted by intron 2 of the γ-globin gene in the sense orientation. The splice donor (SD) and splice acceptor (SA) sites of the intron are indicated on the figure. The *neo* gene is also flanked by a heterologous promoter (P') and a polyadenylation signal (A') denoted by the striped triangles. Transcripts originating from the promoter driving L1.2*mneoI* expression (P) can splice the intron, but continue to contain an antisense copy of the *neo* gene. G418-resistant ($G418^R$) colonies should arise only when this transcript is reverse transcribed, integrated into chromosomal DNA, and expressed from its own promoter, P'.

Figure 2A is a diagram depicting cloning of *L1.2mneoI. L1.2mneoI* cloned into pCEP4 to create pJM101. pCEP4 contains an origin of replication (Ori) and a selectable marker (*Amp*) for prokaryotic cells and an origin of replication and transacting factor (Ori/EBNA1) and a selectable marker (*Hyg*) for eukaryotic cells. The direction of transcription of each gene is denoted by arrows. The features of *L1.2mneoI* are described in the description of Figure 1.

Figure 2B is a diagram depicting mutant constructs of *L1.2mneoI.* pJM102 lacks the 910 bp 5' UTR of L1.2; pJM103 has a 3.8 kb deletion wherein most of the 5' UTR, all of ORF1 and the first 2.1 kb of ORF2 are deleted; pJM105 contains a missense mutation (D702Y) in ORF2. Each of the mutants have the pCEP4 sequences as the vector portion.

Figure 3A is a diagram outlining the *L1.2mneoI* retrotransposon assay. HeLa cells were transfected with the desired constructs using lipofectamine. Hygromycin-resistant

(hyg^R) cells expressing the wild type and mutant constructs of *L1.2mneoI* were harvested 12-14 days later.

Figure 3B is an image of the results of the retrotransposition assay. $G418^R$ foci were fixed to flasks and stained with Giemsa for visualization. Flasks containing cells transfected with pJM101, pJM102, pJM103 and pJM105 are shown.

Figure 4A is an image of a Southern blot of $G418^R$ clones following retrotransposition with *L1.2mneoI*. Genomic DNA was isolated from four independent $G418^R$ clones (lanes A-D). Approximately 20 μg of each DNA was restricted with EcoRI and was subjected to Southern blot analysis using a 0.46 kb *neo* gene as a probe. The size of the molecular weight standards used is indicated on the figure.

Figure 4B is an image of a gel depicting precise splicing of the intron present in the original retrotransposon construct and an image of the structure of the products. 500 ng of genomic DNA from clones A-D was used as individual templates in PCR reactions using the primers neo437S and neo1808AS. One fifth volume of the products was separated on a 1.0% agarose gel containing ethidium bromide. A 468 bp DNA fragment diagnostic for the loss of the intron was detected in each clone (lanes 2, 3, 4 and 5). In addition, a small amount of a 1361 bp DNA fragment diagnostic for the original vector was observed in lanes 2, 3 and 4. Lane 6 contains DNA from HeLa cells and lane 7 is a DNA negative control. Lane 1 contains a 1 kb molecular weight size ladder (Gibco/BRL).

Figure 5 is a diagram depicting the genomic structures of the insertions A-D. Each insertion was compared to its corresponding 'empty site' which was independently cloned from HeLa cell genomic DNA. Truncated portions of *L1.2mneoI* are shown and the nucleotide position of the truncation in L1.2 is noted. Dark filled rectangles are L1.2 sequences and hatched rectangles are the SV40 promoter and SV40 poly A signal at the two ends of the antisense *neo* gene. Dotted rectangles are transduced sequences between the 3' end of L1.2 and the poly A site derived from the pCEP4 vector. Open rectangles represent genomic DNAs. Rightward arrows indicate target site duplications. The length of the poly A tracts and the sizes of the target site duplications and/or deletions are indicated. The arrow flanking insertion A is marked parenthetically because the target site could be a 1-2 bp duplication, a blunt insertion or up to a 4 bp deletion.

Figure 6 is a diagram depicting mutant constructs of *L1.2mneoI* transfected into HeLa cells. The approximate positions of ORF1, ORF2 and Δ3' UTR mutants are indicated. Each mutant was constructed in the pJM102 backbone and lacks the 5' UTR sequence of L1.2. Wild type amino acids which were mutated are underlined and the resulting mutant sequence is shown below the underline (SEQ ID NO:138-152).

Figure 7 is a diagram depicting the sequence of various poly A elements and human AP endonuclease. The structure of the human L1 element is also shown wherein PROM denotes the L1 internal promoter; vTSD denotes the variable target site duplication; EN denotes the endonuclease domain; RT denotes the reverse transcriptase domain; and, ZN denotes the putative Zn-finger-like domain. The amino acid sequence alignment of poly A elements and human AP endonuclease are shown wherein the sequences are: TAD, from *Neurospora crassa* (SEQ ID NO:1 through 7); L1Tc, from *T. cruzi* (SEQ ID NO:8-14); R1Bm, from *B. mori* (SEQ ID NO:15-21); FDM and GDM (F and G elements) from *D. melanogaster* (SEQ ID NO:22-35); IDM (I-factor) from *D. Teissieri* (SEQ ID NO:43-49); Jock, *jockey* from *D. melanogaster* (SEQ ID NO:36-42); L1Hs, human L1 (SEQ ID NO:50-56; Tx1, from *Xenopus laevis* (SEQ ID NO:57-63), Cin4, from *Zea mays* (SEQ ID NO:64-70); and DRE, from *Dictyostelium discoideum* (SEQ ID NO:71-77). APHs is the human AP endonuclease (SEQ ID NO:78-84), DNase I from bovine pancreas (SEQ ID NO:85-89). The EN domain was also identified in the following elements: CR1 (chicken), *ingi* (trypanosome), L1Md (mouse, and other mammalian L1s), Ta11 (*Arabidopsis*), TART (*D. melanogaster*), TRAS (*B. mori*), T1 (mosquito). Conserved (>2 identities) residues are shaded; residues conserved among all poly A elements and the human AP endonuclease are represented by a single circle; putative active site residues are indicated by a double circle. The numbers refer to the residues between two conserved blocks. Residues mutated in L1 ENp are indicated by arrows and the names of each of the mutants are shown below each of the mutations.

Figure 8 is an image of a series of gels depicting purification of and nicking activities of L1 ENp and mutant proteins. In the gel labeled (A), purified proteins were separated on a 10% SDS-PAGE gel and were stained with Coomassie Blue. Approximately equal amounts of protein were loaded except in the case of H230A wherein 10-fold less protein was loaded. MW, molecular weight standards. In the gel labeled (B) the nicking activities of the proteins were assessed. The lanes are numbered left to right and contain the following: 1) phage

λ Hind III digest MW marker; 2) substrate pBS DNA, no protein added; 3) with 2.6 ng wild-type L1 ENp; 4) with 26 ng wild-type L1 ENp; 5) E43A mutant; 6) D205G; 7) N14A; 8) D145A; 9) H230A. The symbols used are as follows: sc is supercoiled plasmid; oc is open (nicked) circular plasmid; l is linear plasmid. In the gel labeled C. nicking was examined over time. Essentially, 50 fmol L1 ENp (or D205G mutant) was used to digest 500 fmol pBS and the extent of nicking was measured at the indicated times.

Figure 9 is an image of a gel depicting the structure of the nicked DNA and preference of the enzyme for a supercoiled substrate. Supercoiled pBS DNA (0.2 μg) (lane 2) was incubated with L1 ENp to generate open circle DNA (lane 3). Subsequently L1 ENp was heat inactivated, and T4 DNA ligase was added (lane 4). After ligation, T4 DNA ligase was heat inactivated, and the product was again incubated with L1 ENp (lane 5). Lanes 7-10 are similar, except that 10-fold less L1 ENp was added initially. The symbol cc denotes closed relaxed circle DNA.

Figure 10 is an image of a gel depicting the fact that L1 ENp cleaves native DNA and apurinic DNA equally well. The DNA substrate was either native DNA or apurinic DNA. KS-DNA, native pBS KS(-) DNA; AP-DNA, apurinic DNA; sc, supercoiled DNA; oc, open circle DNA, MW, λ HindIII digest.

Figure 11 is a series of gels and a sequence depicting cleavage hotspots in pBS plasmid. In the gel labeled (A), L1 ENp double-strand break hotspot is shown. Linear pBS DNA products were electroeluted, digested with restriction enzymes, and run on agarose gels. The gel labeled (B) depicts the L1 ENp cleavage reaction. Lane 1, supercoiled DNA substrate; lanes 2-5, 13 ng, 26 ng, 65 ng and 130 ng of L1 ENp added to 3.2 μg DNA, respectively; 5% of these samples were run on the gel. In the gel labeled (C), primer extension on uncleaved substrate and L1 ENp products was performed on the products shown in (B). A sequence ladder generated with the indicated kinased primer was included for each reaction. Primers JB1132 and JB1133 are specific for each strand flanking the cleavage hotspot region of pBS. In the sequence labeled (D; SEQ ID NO:90), cleavage hotspots in pBS are shown. Major cleavage sites are denoted by large vertical arrows; minor cleavage sites are denoted by smaller vertical arrows; horizontal arrows indicate inverted repeats (heavy arrows, pBR322 minor; thin arrows, pBR322 sub-minor; Lilley, 1981, *Nucl. Acids Res.* 9:1271-1288).

Figure 12 is an image of a gel depicting cleavage specificity of the enzyme, which cleavage does not require supercoiling. DNAs were treated with L1 ENp and used as templates for primer extension experiments as in Figure 11. Lanes 1, supercoiled DNA, no L1 ENp; lanes 2, supercoiled DNA + 20 ng L1 ENp; lanes 3, relaxed closed circular DNA, no L1 ENp; lanes 4, relaxed closed circular DNA + 80 ng L1 ENp. GATC lanes indicate sequencing reactions primed with the indicated kinased oligonucleotide.

Figure 13 (SEQ ID NO:91) is a sequence diagram depicting the fact that K-DNA contains a hotspot for L1 ENp cleavage (indicated by bold arrow). The cleavage sites were determined as described in Figure 11 except that the SP6 primer was used. Sites of enhanced cleavage by hydroxyl radical were determined using the method of Burkhoff *et al.* (1987, *Cell* 48:935-943) and are indicated by small vertical arrows. Bold letters indicate phased A-tracts.

Figure 14 is a diagram of a series of sequences depicting the similarity of *in vitro* cleavage sites for L1 ENp and the predicated sites of priming of reverse transcription. In the diagram labeled (A; SEQ ID NO:92-94), a model based on the JH-25 sequence for concerted target DNA nicking and reverse transcription of the 3' poly A end of L1 RNA is shown. The specificity of L1 ENp for $(Py)_n \downarrow (Pu)_n$ generates a polypyrimidine 3' terminus that can in principle base pair to the 3' poly A of L1 RNA. Such complementarity might stabilize a reverse transcription priming complex (B-G). Comparison of cleavage sites determined *in vitro* (shown in B; SEQ ID NO:95-106) to various *in vivo* inferred priming sites involved in L1 retrotransposition is also shown. Note that the nucleotide 3' to the cleavage site is always a purine, is usually an A, and is usually part of an oligopurine run (boxed residues). In many cases, there is a symmetrically placed oligopyrimidine tract 5' to the cleavage site or inferred priming site (underlined residues). For parts (C-G) letters in lower case represent the TSD. Note that the runs of As at the 5' end of many of the TSDs represent an area of microhomology with the 3' poly A tract of the L1 insertion. These are assumed to represent part of the TSD here. In the diagram labeled (B), pBS targets are shown. The top strand is arbitrarily defined as the strand cleaved first. In the diagram labeled (C; SEQ ID NO:107-111), new mutations caused by L1 insertion are shown. These include three hemophilia A mutations (Kazazian et al., 1988, Nature 332:164-166; Woods-Samuels et al., 1989, Genomics 4:290-296) and a dystrophin mutation (Holmes et al., 1994, Nature Genet. 7:143-148), and a somatic insertion into the APC tumor suppressor gene associated with cancer (Miki et al., 1992, Cancer Res. 52:643-645). In

the diagram labeled (D; SEQ ID NO:112-114), new L1-*neo* transposition events that occur in HeLa cells and described herein are shown. In the diagram labeled (E;SEQ ID NO:115-116), active transposon copies discovered as progenitor elements for the JH-27 insertion (L1.2) and the dystrophin insertion (LRE2) are shown. In the diagram labeled (F; SEQ ID NO:117-121), other full length elements cloned intentionally in searches to find active elements L1.1-L1.4 (Dombroski *et al.*, 1991, *Science* 254:1805-1808; Dombrowski *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:6513-6517), CGL1.1 (Hohjoh *et al.*, 1990, *Nucl. Acids Res.* 18:4099-4104) or discovered by searching for element copies in GenBank (Z73497) are shown. In the diagram labeled (G; SEQ ID NO:122-129), GenBank was searched using BLASTN with the 3' UTR sequence of L1.2 and the top 34 hits were studied. Approximately half of the truncated elements had a precise TSD. These are all listed on this Figure, identified by the appropriate GenBank accession number.

Figure 15 is a diagram and a table showing that the L1 En domain is required for transposition in HeLa cells. In (A), a diagram of the L1.2*mneoI* retrotransposition assay is shown. A *neo* marker gene with a "backward" intron (*mneoI*) is inserted upstream of L1 3' UTR such that *neo* and L1 are convergently transcribed. L1 transcription from the CMV promoter leads to the splicing of the intron and reconstruction of the *neo* coding region. Reverse transcription and integration leads to expression of *neo* from its SV40 promoter, pCMV, cytomegalovirus early promoter; S.D., splicing donor; S.A., splicing acceptor; wavy line, RNA; V, intron sequence. In (B), L1 retrotransposition frequencies are tabulated. D703Y is the RT active site mutant; the other mutants are EN domain mutants.

Figure 16 is a diagram depicting the organization of a human L1 element and the location of oligomers A, B and C. ORF1 and ORF2 are indicated by a light gray box and dark gray box, respectively. The 5' and 3' untranslated regions (UTRs) are indicated by striped boxes and the poly A tail by A_n. The approximate positions of the endonuclease (EN), reverse transcriptase (RT) and cysteine-rich motifs in ORF2 are indicated. Oligomer A is located at nucleotides 61-80, oligomer B at nucleotides 941-960, and oligomer C at nucleotides 5919-5938 of the L1.2 sequence (Dombroski, et al., 1991, *Science* 254:1805-1808).

Figure 17A is a diagram depicting the Ty1-based construct used to express the L1 RT in the biochemical assay shown in Figure 17B. Ty1 contains two ORFs. The first, TyA, encodes a Gag-like protein. The second TyB, is expressed as a fusion protein that is post-

translationally processed to generate proteins with protease, integrase, RT, and RNase H activity. When Ty1 is experimentally expressed from a promoter inducible by galactose (GAL1), the Ty1-encoded proteins and RNA co-assemble into cytoplasmic virus-like particles (VLPs) which can be partially purified and assayed for Ty1 RT activity (Garfinkel et al., 1985, *Cell* 42: 507-517). The integrase, RT and RNase H domains of TyB are replaced by L1 ORF2. The hemagglutinin epitope tag 12CA5 (et) was inserted at the Ty1/L1 ORF2 junction. Boxes with black triangles are long terminal repeats (LTRs). Expression from the inducible *GAL1* promoter results in virus-like particles (VLPs) that contain RT.

Figure 17B is a graph depicting the RT activity of thirteen novel L1 elements and L1.3, L1.4, and LRE2. One mg of total VLP extract in a 30 ml reaction volume was assayed as described in the materials and methods section of Example 3. Relative RT activity is reported as fmoles of a^{32}P-dGTP incorporated into a polyrC/oligodG template. Values are the averages of 5-8 independent assays of two separate VLP preparations and the error bars are shown. RT activity a levels significantly greater than that observed for the D702Y mutant was observed in the case of L1.3, L1.4, L1.6, L1.12, L1.15, L1.19, L1.21, L1.25, and L1.33.

Figure 17C is a graph depicting the results of the *HIS3* pseudogene assay. Constructs containing the reverse transcriptase domain of each L1 element were transformed into yeast strain YDS50.1. His+ prototroph formation requires the presence of a functional reverse transcriptase. The frequency of positive events was determined for at least eight independent transformants derived from at least two separate experiments. The substantial range of frequencies observed necessitated the production of a high-range frequency graph, in which LRE1 serves as a positive control, as well as a low-range graph, in which LRE2 serves as a positive control.

Figure 18A is a diagram of an overview of the L1 retrotransposon system in Example 3. L1.2 was tagged with an indicator gene (*mneoI*) designed to detect retrotransposition events as described herein. The indicator gene contains an antisense copy of the *neo* gene disrupted by intron 2 of the g-globin gene in the sense orientation (Freeman et al., 1994, *BioTechniques* 17: 47-52). The splice donor and acceptor sites of the intron are indicated. The *neo* gene is also flanked by heterologous polyadenylation (A') and promoter (P') sequences denoted by the hatched rectangles. Transcripts originating from the promoter driving L1 expression (P) are spliced, but contain a non-functional copy of the *neo* gene. G418 resistant

($G418^R$) cells arise only when the L1 mRNA is reverse transcribed, integrated into HeLa chromosomal DNA, and expressed from its own promoter (P').

Figure 18B is a series of images depicting the retrotransposition frequency of various L1 elements. A one hundred-fold variation in retrotransposition frequency among active L1 elements. The retrotransposition assay was performed as described herein. $G418^R$ cells were fixed to flasks and stained with Giemsa for visualization. Flasks containing cells transfected with L1.2, D702Y, L1.3, L1.4, L1.19, L1.20 and L1.39 are shown.